Semiparametric Transformation Models for Semicompeting Survival Data

Huazhen Lin* Ling Zhou†
Chunhong Li‡ Yi Li**

*Southwestern University of Finance and Economics
†Southwestern University of Finance and Economics
‡Sichuan University
**University of Michigan - Ann Arbor, yili@umich.edu

This working paper is hosted by The Berkeley Electronic Press (bepress) and may not be commercially reproduced without the permission of the copyright holder.
http://biostats.bepress.com/umichbiostat/paper107
Copyright ©2013 by the authors.
Semiparametric Transformation Models for Semicompeting Survival Data

Huazhen Lin, Ling Zhou, Chunhong Li, and Yi Li

Abstract

Semicompeting risk outcome data, e.g. time to disease progression and time to death, are commonly collected in clinical trials, but complicated analytical tools hamper the analysis and the interpretation of the results. We propose a novel semiparametric transformation model for such data. Compared with the existing models, our model is advantageous in the following distinctive ways. First, it allows us to provide direct estimators of the regression analysis and the association parameter. Second, the measure of surrogacy, for example, the proportion of treatment effect and relative effect, can also be directly obtained. We propose a two-stage estimation procedure for inference and show the proposed estimator is consistent and asymptotically normal. Extensive simulations demonstrate the valid usage of our method. We apply the method to a real cancer trial to study the impact of several biomarkers on patients’ semicompeting outcomes, namely, time to progression and time to death.
Semiparametric transformation models for semicompeting survival data

Huazhen Lin*, Ling Zhou*, Chunhong Li † and Yi Li‡

SUMMARY Semicompeting risk outcome data, e.g. time to disease progression and time to death, are commonly collected in clinical trials, but complicated analytical tools hamper the analysis and the interpretation of the results. We propose a novel semiparametric transformation model for such data. Compared with the existing models, our model is advantageous in the following distinctive ways. First, it allows us to provide direct estimators of the regression analysis and the association parameter. Second, the measure of surrogacy, for example, the proportion of treatment effect and relative effect, can also be directly obtained. We propose a two-stage estimation procedure for inference and show the proposed estimator is consistent and asymptotically normal. Extensive simulations demonstrate the valid usage of our method. We apply the method to a real cancer trial to study the impact of several biomarkers on patients’ semicompeting outcomes, namely, time to progression and time to death.

KEY WORDS: Semicompeting risk data; Semiparametric linear transformation model; Surrogate endpoints; Two-stage estimation.

1 Introduction

Terminal events such as death are often the main endpoint of clinical trials on patients with chronic life-threatening diseases, e.g. cancer. In the evolving course of the disease, landmark events, for example disease progression, are also observed. Such non-terminal events are typically the precursors of the main event and also serve as important endpoints in clinical trials. It is often of substantial interest to study
(1) the association between the landmark event and the death, and (2) the marginal
distribution of the time to the landmark event and the time to death given treatment
and other underlying individual characteristics.

Denote the time to landmark event by $S$ and the time to death by $T$. Given
that the occurrence of the terminal events precludes the occurrence of the non-terminal
events, but not vice versa, $S$ and $T$ fall into the paradigm of semi-competing risk
data (Fine, Jiang, and Chappell, 2001). A variety of methods have been proposed
to model $S$ and $T$. For example, Day, Bryant, and Lefkopolou (1997) considered
the Clayton-Oakes model (Clayton, 1978; Oakes, 1986) and proposed a test of the
independence of $T$ and $S$. Fine et al. (2001) provided a closed form estimator
of the association parameter in the Clayton-Oakes model using modified weighted
concordance estimating functions from Oakes (1982, 1986) along with an asymptotic
variance estimator. Wang (2003) proposed an estimation procedure in this model
which is applicable more generally to copula models.

In the aforementioned works, the dependence between the landmark event and the
death is assessed marginally with no adjustment for covariates such as sex, age, or
treatment group made. In practice, the distributions of $T$ and $S$ in the subpopulations
defined by treatment, sex, or age are considered. Regression methodology offers
an opportunity to investigate how patient characteristics influence landmark event
and death. The literature on regression analysis tailored to semicompeting risks
is limited. Lin et al. (1996) introduced a semi-parametric bivariate location-shift
model to describe the effect of treatment on landmark event and death in two-arm
randomized studies. The model can be written as follows:

\begin{align}
H(S) &= X\beta + \varepsilon_1, \\
H(T) &= X\alpha + \varepsilon_2,
\end{align}

where $H(x) = \log(x)$, $\beta$ and $\alpha$ are parameter scales, $(\varepsilon_1, \varepsilon_2)'$ are correlated error terms
with unspecified distribution, and sole covariate $X$ is the treatment indicator. Chang
(2000) extended Lin et al.’s method to the semi-competing risk data with general
discrete covariate. This research direction has been further extended to general re-
gression settings in which the non-terminal event is generalized to be recurrent events
(Ghosh and Lin, 2003; Lin and Ying, 2003), whereas death still serves as a terminal
event. Recently, Ghosh (2009) applied Lin et al.’s model to assess surrogacy. It is
difficult to extend Lin et al. and Chang’s method to high-dimensional discrete co-
variate or continuous covariate because excessive artificial censoring can occur. Lin et
al. and Chang’s methods work for a low-dimensional discrete covariate. In addition, because the distributions of the error terms are completely unspecified, Lin et al. and Chang’s methods cannot make estimation and inference on the association between $S$ and $T$ based on the bivariate location-shift models (1.1) and (1.2); extra steps or models are required to obtain the association parameter. Recently, to consider both the marginal effect of covariates on the landmark event and the association between $S$ and $T$, Hsieh, Wang and Ding (2008) considered a method that combined the copula model and the model (1.1) with either $H$ or the distribution of error is known. However, the methodology proposed by Hsieh, Wang and Ding (2008) was developed for discrete covariates again.

When analyzing non-standard data like survival data, an investigator has to consider where to place assumptions and where to keep the model flexible. The methods proposed by Lin et al. (1996) and Chang (2000) allowed the distributions of the error terms to be unknown, but required the specifications of the transformation functions. Hsieh, Wang and Ding’s method provide investigators an opportunity to place an assumption on the transformation function or the distribution of error. However, all these methods required an extra model for the association.

In the present paper, a new approach is adopted. Our model not only directly provides the marginal regression models of $S$ and $T$, but also the association parameter between $S$ and $T$. To illustrate our idea, we consider the case without covariates. Denote the distributions of $S$, $T$ and the standard normal variable by $F_1$, $F_2$ and $\Phi$, respectively, hence $F_1(S)$ and $F_2(T)$ marginally follow a uniform distribution on $[0, 1]$, and its probit-type transformation $\Phi^{-1}(F_1(S)) \equiv H_1(S)$ and $\Phi^{-1}(F_2(T)) \equiv H_2(T)$ follow the standard normal distribution marginally. Correlation between $H_1(S)$ and $H_2(T)$ within the traditional Gaussian framework is then imposed conventionally and leads to the normal copula model (Li and Lin, 2006). With the covariates in mind, in this paper, we consider the following models,

$$
H_1(S) = X'\beta + \varepsilon_1, \quad (1.3) \\
H_2(T) = X'\alpha + \varepsilon_2, \quad (1.4)
$$

where $H_1$ and $H_2$ are unknown monotonic increasing transformation functions, $

\begin{pmatrix}
\varepsilon_1 \\
\varepsilon_2
\end{pmatrix}
\sim N(0, \Sigma_{\rho}), \quad \Sigma_{\rho} = \begin{pmatrix} 1 & \rho \\
\rho & 1 \end{pmatrix}.

Here, assume $Var(\varepsilon_2) = Var(\varepsilon_1) = 1$ and that $X$ excludes the intercept term for
the identification of the models. $X$ can be continuous covariate, discrete covariate, or combination of continuous and discrete covariates. The models (1.3) and (1.4) leave the transformation functions unspecified, but require the error distribution to be Gaussian. Three-fold reasons for this. One is the transformation function is more fundamental than the error distribution in estimating the regression coefficients (Lin and Zhou, 2009). Specifically, the misspecification of the transformation function leads to a seriously biased estimator of the regression coefficients, while the misspecification of the error distribution leads to a slightly biased or essentially unbiased estimator. Secondly, the use of Gaussian error provides an opportunity to model the association between $S$ and $T$. Finally, the normal distribution is robust in some degrees (Hanley, 1988). The models (1.3) and (1.4) naturally provide not only the marginal regression models of $S$ and $T$, but also the association parameter of $S$ and $T$. Contrarily, the models proposed by Lin et al. (1996) and Chang (2000) cannot provide the direct association parameter, while Hsieh, Wang and Ding (2008) requires an extra copula model for the association parameter.

The remainder of the article is organized as follows. A two-stage estimation procedure is described in Section 2. The asymptotic properties are derived in Section 3. Section 4 contains the simulation results and an application to ?? study. Section 5 gives some concluding remarks.

2 A two-stage estimation procedure

We start by making the following definitions. Let $a \land b = \min(a, b)$ and $I(A)$ be the indicator function for the event $A$. Let $C$ be the time to censoring and $X$ the $p$-dimensional covariate vector. Assume that $(S, T)$ and $C$ are conditionally independent given $X$. We have $n$ observations $(U_{1i}, \delta_{1i}, U_{2i}, \delta_{2i}, X_i), i = 1, \cdots, n$, a random sample from $(U_1, \delta_1, U_2, \delta_2, X)$, where $U_1 = S \land T \land C$, $\delta_1 = I(S \leq T \land C)$, $U_2 = T \land C$, and $\delta_2 = I(T \leq C)$. Hence, $S$ is censored by the minimum of $T$ and $C$ and not just by $C$. The dependent censoring will complicate the analysis. For notational simplicity, denote the parameter vectors $\beta, \alpha$ and $\rho$ by $\Theta$. Hence, $\Theta$, $H_1$ and $H_2$ are the parameters and nonparametric functions defined by the models (1.3) and (1.4).

2.1 Estimation of the parameters

Since the distribution of $(\varepsilon_1, \varepsilon_2)'$ is known, the parameters and the transformation
functions can be estimated by the maximum likelihood function. For each observation
$i$, the likelihood will take one of four forms defined below depending on the values of
$\delta_{1i}$ and $\delta_{2i}$.

(1) If both events are observed ($\delta_{1i} = 1, \delta_{2i} = 1$),

$$L_{i1}(\Theta; H_1, H_2) \propto \frac{\phi(H_1(U_{1i}) - X_i'\beta)}{\sqrt{1 - \rho^2}} \phi \left( \frac{H_2(U_{2i}) - X_i'\alpha - \rho(H_1(U_{1i}) - X_i'\beta)}{\sqrt{1 - \rho^2}} \right),$$

where $\phi$ is the density function of the standard normal random variable.

(2) If $S_i$ is observed, but $T_i$ is not observed ($\delta_{1i} = 1, \delta_{2i} = 0$),

$$L_{i2}(\Theta; H_1, H_2) \propto \phi(H_1(U_{1i}) - X_i'\beta) \left\{ 1 - \Phi \left( \frac{H_2(U_{2i}) - X_i'\alpha - \rho(H_1(U_{1i}) - X_i'\beta)}{\sqrt{1 - \rho^2}} \right) \right\},$$

(3) If $S_i$ is not observed, but $T_i$ is observed ($\delta_{1i} = 0, \delta_{2i} = 1$),

$$L_{i3}(\Theta; H_1, H_2) \propto \phi(H_2(U_{2i}) - X_i'\alpha) \left\{ 1 - \Phi \left( \frac{H_1(U_{1i}) - X_i'\beta - \rho(H_2(U_{2i}) - X_i'\alpha)}{\sqrt{1 - \rho^2}} \right) \right\},$$

(4) If neither event is observed ($\delta_{1i} = 0, \delta_{2i} = 0$),

$$L_{i4}(\Theta; H_1, H_2) \propto \int_{H_1(U_{1i})}^{\infty} \int_{H_2(U_{2i})}^{\infty} \phi(x) \phi \left( \frac{y - \rho x}{\sqrt{1 - \rho^2}} \right) dxdy.$$

Combining these, the likelihood resulting from observation $i$ yields

$$L_i(\Theta; H_1, H_2) \propto L_{i1}(\Theta; H_1, H_2)^{\delta_{1i}} L_{i2}(\Theta; H_1, H_2)^{\delta_{2i}} L_{i3}(\Theta; H_1, H_2)^{1-\delta_{1i}} \times L_{i4}(\Theta; H_1, H_2)^{1-\delta_{2i}}.$$

The likelihood function involves both finite dimensional parameters $\Theta$ and infinite
dimensional parameters $H_1$ and $H_2$. Maximization of the likelihood function over
an infinite dimensional parameter space can be complicated, especially when the
objective function involves two unknown functions $H_1$ and $H_2$. In fact, even for the
simple case involving one unknown transformation function for single survival data
with independent censoring and without semicompeeting, to obtain nonparametric
maximum likelihood estimator of the transformation, Zeng and Lin (2006) proposed a
quasi-Newton method with an search along gradients of the loglikelihood function, and
in each iteration of the search, a large linear system is required to be approximately solved by using the method of preconditioned conjugate gradients (Zeng and Lin, 2006). Because the focus of the paper is the estimation of $\Theta$, in the current paper, we propose a two-stage approach to estimate $\Theta$, $H_1$ and $H_2$. Our method is considerably easy to computation of the estimators for $H_1$ and $H_2$, while the loss of efficiency on the estimators for $\Theta$ is little because $\Theta$ is estimated by maximizing a pseudo-likelihood. Particularly, a series of estimating equations described in Section 2.2 is used to estimate the transformation functions $H_1$ and $H_2$. Then, the parameter $\Theta$ is estimated by maximizing a pseudo-likelihood, which is the likelihood function $\prod_{i=1}^{n} L_i(\Theta; H_1, H_2)$ with $H_1$ and $H_2$ replaced by the estimated values. We repeat the procedures of estimating $\Theta$ and $H_1$, $H_2$ until convergence.

2.2 Estimation of the transformation functions given $\Theta$

Both (1.3) and (1.4) are members of the family of semiparametric transformation models and they generalize the well-known Box-Cox transformation model whose transformation function is parameterized by the family of power functions. The model (1.3) (or (1.4)) is also an alternative to the Cox regression model and the proportional odds model, in which $\varepsilon_1$ (or $\varepsilon_2$) is assumed to follow the extreme value distribution (or logistic distribution) instead of being normally distributed. Statistical inference procedures on the single semiparametric transformation model with independent censoring have been developed by Dabrowska and Doksum (1988), Cheng et al. (1995), Chen et al. (2002), and Zeng and Lin (2006) among others. Hence, the consistent estimator of $H_2$ is available due to the independence of $T$ and $C$ given $X$. Here, we use the method proposed by Chen et al. (2002), which is easy to compute. Denote $\alpha_0$ and $H_{20}$ to be the true values of $\alpha$ and $H_2$, respectively, and $\Lambda(t) = -\log(1 - \Phi(t))$ to be the cumulative hazard function of $\varepsilon_2$. Suppose

$$N_2(t) = \delta_2 I(U_2 \leq t), \quad Y_2(t) = I(U_2 > t).$$

Motivated by the fact that $M_2(t) = N_2(t) - \int_0^t Y_2(s)d\Lambda(H_{20}(s) - X'_i\alpha_0)$ is a martingale process, we estimate $H_2(t)$ by the following estimating equation:

$$\sum_{i=1}^{n} \{dN_2(t) + Y_2(t)d\log (1 - \Phi(H_2(t) - X'_i\alpha))\} = 0, \quad (2.2)$$

where $H_2$ satisfies $H_2(0) = -\infty$. The requirement ensures that $\Lambda(a + H_2(0)) = 0$ for any finite $a$. It is easy to see that the estimator of $H_2$ is a nondecreasing step function
on $[0, \infty)$ with $H_2(0) = -\infty$ and with jumps only at the observed uncensored terminal event times, denoted by $t_{d,1} < \cdots < t_{d,K}$.

Now, consider the estimation of $H_1$. Since $S$ and $T$ are correlated, the direct uses of Chen et al.'s methods on the data \{(U_{1i}, \delta_{ii}), i = 1, \cdots, n\} would yield an inconsistent estimator of $H_1$ due to dependent censoring. Alternatively, using the idea of Lin and Ying (1993) and Hsieh, Wang and Ding (2008), one can estimate $H_1(t)$ based on the identity

$$EI(U_{1i} \geq t, U_{2i} \geq t) = S_{\rho}(H_1(t) - \mathbf{X}_i'\beta, H_2(t) - \mathbf{X}_i'\alpha) \Pr(C_i > t|\mathbf{X}_i),$$

where $S_{\rho}$ is the survival function of $N(0, \Sigma_{\rho})$. A problem of the method is that the distribution of the censoring time $C$ requires modelling.

In the paper, a different approach is used. Our approach does not involve the distribution of $C$. A key observation to obtain our estimator is that

$$S_i \wedge (T_i \wedge C_i) = (S_i \wedge T_i) \wedge C_i,$$

which implies that the survival analysis in which $S_i$ is the survival time and $T_i \wedge C_i$ is the censoring time, can be regarded as the survival analysis in which $S_i \wedge T_i$ is the survival time and $C_i$ is the censoring time. Given that $\mathbf{X}_i$, $(S_i, T_i)$ is independent of $C_i$, by regarding the survival time as $W_i = S_i \wedge T_i$ and the censoring time as $C_i$, we obtain an independent censoring problem. Then, the use of the method proposed by Chen et al. (2002) to the data \{(W_i \wedge C_i, I(W_i \leq C_i), \mathbf{X}_i) : i = 1, \cdots, n\} would yield consistent estimators of related parameters and functions. Under the models (1.3) and (1.4), $H_1$ and $H_2$ are monotonic increasing functions, for any $t$, we get

$$P(W \geq t|\mathbf{X}) = P(S \wedge T \geq t|\mathbf{X}) = P(S \geq t, T \geq t|\mathbf{X}) = P(H_1(S) \geq H_1(t), H_2(T) \geq H_2(t)|\mathbf{X}) = S_{\rho}^H(H_1(t) - \mathbf{X}'\beta, H_2(t) - \mathbf{X}'\alpha),$$

hence, the cumulative hazard function of $W$ is given by

$$\hat{\Lambda}(t) = -\log \left\{ S_{\rho}(H_1(t) - \mathbf{X}'\beta, H_2(t) - \mathbf{X}'\alpha) \right\}.$$

Denote

$$N_i(t) = \eta_i I(U_{1i} \leq t), \quad \eta_i = I(W_i \leq C_i) \quad \text{and} \quad Y_i(t) = I(U_{1i} \geq t),$$

motivated by

$$M_i(t) = N_i(t) + \int_0^t Y_i(s) d\log \left\{ S_{\rho_0}(H_{10}(s) - \mathbf{X}_{10}'\beta_0, H_{20}(s) - \mathbf{X}_{20}'\alpha_0) \right\}$$


is a martingale process, given $\Theta$ and $H_2$, we estimate $H_1(t)$ by the following equation:

$$
\sum_{i=1}^{n} \{dN_i(t) + Y_i(t) d \log \{S_\rho (H_1(t) - \mathbf{X}'i\beta, H_2(t) - \mathbf{X}'i\alpha)\}\} = 0, \quad (2.3)
$$

where $H_1(0) = -\infty$. Again, following the estimating equation (2.3), the estimator $\hat{H}_1(\cdot)$ of $H_1(\cdot)$ is a step function with jumps at a combination of the observed censored terminal and non-terminal event time, denoted by $t_1 < \cdots < t_M$. Thus solving the system of estimating equations of infinite number of equations defined by (2.2) and (2.3) is equivalent to solving the system of finite number of equations. In addition, unlike a traditional nonparametric approach to estimate the transformation function (Horowitz, 1996; Zhou, Lin and Johnson, 2009), our approach does not involve non-parametric smoothing, and thus does not suffer from smoothing-related problems, for example, selection of a smoothing parameter. Finally, the estimator of $H_2$ is independent of the estimator of $H_1$, hence the estimations of the two infinite-dimensional parameters is decomposed into two separate estimations of single infinite-dimensional parameters, which can greatly reduce computational cost.

2.3 Algorithm to estimate $\Theta$, $H_1$ and $H_2$

Using the idea of Chen et al. (2002), for easy computation, alternative versions of (2.2) and (2.3) are provided. Note that (2.2) and (2.3) can be rewritten as

$$
\sum_{i=1}^{n} \left\{dN_i(t) + Y_i(t) \left[\log (1 - \Phi (H_2(t) - \mathbf{X}'i\alpha)) - \log (1 - \Phi (H_2(t-) - \mathbf{X}'i\alpha))\right]\right\} = 0, \quad (2.4)
$$

$$
\sum_{i=1}^{n} \left\{dN_i(t) + Y_i(t) \left[\log \{S_\rho (H_1(t) - \mathbf{X}'i\beta, H_2(t) - \mathbf{X}'i\alpha)\}\right]
- \log \{S_\rho (H_1(t-) - \mathbf{X}'i\beta, H_2(t-) - \mathbf{X}'i\alpha)\}\left\} = 0, \quad (2.5)
$$

with $H_1(0) = H_2(0) = -\infty$. Slightly differently from (2.4) and (2.5), one might also consider the following computationally simpler estimating equations:

$$
\sum_{i=1}^{n} \left\{dN_i(t) - \frac{Y_{2i}(t) \phi (H_2(t-) - \mathbf{X}'i\alpha)}{1 - \Phi (H_2(t-) - \mathbf{X}'i\alpha)} dH_2(t)\right\} = 0, \quad (2.6)
$$

$$
\sum_{i=1}^{n} \left\{dN_i(t) + \frac{Y_i(t)}{S_\rho (H_1(t-) - \mathbf{X}'i\beta, H_2(t-) - \mathbf{X}'i\alpha)} \left[S_\rho^{(10)} (H_1(t-) - \mathbf{X}'i\beta, H_2(t-) - \mathbf{X}'i\alpha) \times dH_1(t) + S_\rho^{(01)} (H_1(t-) - \mathbf{X}'i\beta, H_2(t-) - \mathbf{X}'i\alpha) dH_2(t)\right]\right\} = 0, \quad (2.7)
$$
where \( S_p^{(10)}(x, y) = \partial S_p(x, y)/\partial x \) and \( S_p^{(01)}(x, y) = \partial S_p(x, y)/\partial y \). Equations (2.4), (2.5), (2.6) and (2.7) suggest the following iterative algorithms for computing \( \Theta, H_1 \) and \( H_2 \).

**Step 0.** Choose an initial value of \( \Theta \).

**Step 1.** Obtain \( H_2 \) as follows. First noting that \( H_2(t_{d,1}) = -\infty \) and using (2.4), obtain \( H_2(t_{d,1}) \) by solving

\[
\sum_{i=1}^{n} \{dN_{2i}(t_{d,1}) + Y_{2i}(t_{d,1}) \log (1 - \Phi (H_2(t_{d,1}) - X'(\alpha)))\} = 0.
\]

Then, using (2.6), obtain \( H_2(t_{d,k}), k = 2, \ldots, K \), one-by-one by solving the equations,

\[
H_2(t_{d,k}) = \frac{\sum_{i=1}^{n} dN_{2i}(t_{d,k}) + H_2(t_{d,k-1}) \sum_{i=1}^{n} \frac{Y_{2i}(t_{d,k}) \phi (H_2(t_{d,k-1}) - X'(\alpha))}{1 - \Phi (H_2(t_{d,k-1}) - X'(\alpha))}}{\sum_{i=1}^{n} \frac{Y_{2i}(t_{d,k}) \phi (H_2(t_{d,k-1}) - X'(\alpha))}{1 - \Phi (H_2(t_{d,k-1}) - X'(\alpha))}}.
\]

**Step 2.** Obtain \( H_1 \) as follows. First noting that \( H_1(t_1) = -\infty \) and using (2.5), obtain \( H_1(t_1) \) by solving

\[
\sum_{i=1}^{n} \{dN_i(t_1) + Y_i(t_1) \log (\{S_p(H_1(t_1) - X'(\beta), H_2(t_1) - X'(\alpha))\})\} = 0.
\]

Then, using (2.7), obtain \( H_1(t_k), k = 2, \ldots, M \), one-by-one by solving the equations,

\[
H_1(t_k) = H_1(t_{k-1}) - \frac{\sum_{i=1}^{n} Y_i(t_k) S_p^{(01)}(H_1(t_{k-1}) - X'(\beta) H_2(t_{k-1}) - X'(\alpha)) [H_2(t_k) - H_2(t_{k-1})]}{\sum_{i=1}^{n} Y_i(t_k) S_p^{(10)}(H_1(t_{k-1}) - X'(\beta) H_2(t_{k-1}) - X'(\alpha))},
\]

with \( H_2(t_1), \ldots, H_2(t_M) \) replaced by their estimators obtained in Step 1, noting that \( H_2(t_k) = H_2(t_{k-1}) \) if \( t_k \notin \{t_{d,1}, \ldots, t_{d,K}\} \).

**Step 3.** Obtain the new estimate of \( \Theta \) by maximizing the likelihood \( L_i(\Theta; H_1, H_2) \) defined in (2.1) with \( H_1 \) and \( H_2 \) replaced by the estimators obtained in Steps 1 and 2.

**Step 4.** Set the initial value of \( \Theta \) to be the estimate obtained in Step 3 and repeat Steps 1 to 3 until prescribed convergence criteria are met.
3 Inference in Large Samples

In this section, the large sample properties of all estimators are presented. Let $\hat{\Theta}$, $\hat{H}_1(t)$ and $\hat{H}_2(t)$ denote the estimators of $\Theta$, $H_1(t)$ and $H_2(t)$, and let $\Theta_0$, $H_{10}(t)$ and $H_{20}(t)$ denote the true values of $\Theta$, $H_1(t)$ and $H_2(t)$, respectively. Some notation and regularity conditions are needed. Regularity conditions for ensuring the central limit theorem for counting process martingales such as those assumed in Fleming and Harrington (1991) are assumed here without specific statement. Let $\tau = \inf\{t : P(S_t \wedge T_t > t) = 0\}$. We assume that $\tau$ is finite, $P(S_t \wedge T_t > \tau) > 0$ and $P(C_t = \tau) > 0$. This is to avoid a lengthy technical discussion about the tail behavior. $X_t$ is bounded and $H_{10}$ and $H_{20}$ have continuous and positive derivatives. In the rest of the paper, denote the $(k_1 + k_2 + \cdots)$-th order partial derivative of a function $f(x_1, x_2, \cdots)$ by $f^{(k_1,k_2,\cdots)}(x_1, x_2, \cdots)$; that is, $f^{(k_1,k_2,\cdots)}(x_1, x_2, \cdots) = \frac{\partial^{k_1+k_2+\cdots} f(x_1,x_2,\cdots)}{\partial x_1^{k_1} \partial x_2^{k_2} \cdots}$. Define

$$
\gamma_{1i}(x,y) = -\frac{S_{\rho_0}^{(10)}(x - X_i' \beta_0, y - X_i' \alpha_0)}{S_{\rho_0}(x - X_i' \beta_0, y - X_i' \alpha_0)}; \quad \gamma_{2i}(x,y) = -\frac{S_{\rho_0}^{(11)}(x - X_i' \beta_0, y - X_i' \alpha_0)}{S_{\rho_0}(x - X_i' \beta_0, y - X_i' \alpha_0)},
$$

$$
\gamma_3i(x,y) = -\frac{\hat{S}_\rho(x - X_i' \beta, y - X_i' \alpha)}{\hat{S}_\rho(x - X_i' \beta, y - X_i' \alpha)}, \quad \hat{S}_\rho(x,y) = \frac{\partial S_\rho(x,y)}{\partial \rho},
$$

$$
\gamma_1(H_{10}(t)) = \exp \left( \int_0^t \frac{E[Y_i(s) d\gamma_{1i}(H_{10}(s), H_{20}(s))]}{E[Y_i(s) \gamma_{1i}(H_{10}(s), H_{20}(s))]} ds \right),
$$

$$
\gamma_2(H_{20}(t)) = \exp \left( \int_0^t \frac{E[Y_i(s) d\gamma_{2i}(H_{10}(s), H_{20}(s))]}{E[Y_i(s) \gamma_{2i}(H_{10}(s), H_{20}(s))]} ds \right),
$$

$$
\lambda_{2i}(x) = \frac{\phi(x - X_i' \alpha_0)}{1 - \Phi(x - X_i' \alpha_0)}; \quad \lambda_2(H_{20}(t)) = \exp \left( \int_0^t \frac{E[Y_i(s) d\lambda_{2i}(H_{20}(s))]}{E[Y_i(s) \lambda_{2i}(H_{20}(s))]} ds \right),
$$

$$
\mu(t) = \int_0^t \frac{\lambda_2(H_{20}(s)) E \left[ Y_{2i} \frac{dX_i' \alpha_0}{du} d\lambda_2(H_{20}(s)) \right]}{E[Y_i(s) \lambda_2(H_{20}(s))]}, \quad K(s) = E[Y_i(s) \lambda_2(H_{20}(s))],
$$

$$
B(s) = E[Y_i(s) \gamma_{1i}(H_{10}(s), H_{20}(s))], \quad A(s) = E[Y_i(s) \gamma_{2i}(H_{10}(s), H_{20}(s))],
$$

$$
D_1(s) = E \left\{ \frac{\partial^2 \log L_0(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial H_{10}(U_{1i})} \gamma_1(H_{10}(U_{1i})) \right\} \gamma_1(H_{10}(s)) \left\{ \frac{Y_i(s)}{B(s)} \right\}.
$$
\[ D_2(s) = E \left\{ \frac{\partial^2 \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial H_2(U_{2i})} \frac{Y_{2i}(s)}{\lambda_2(H_{20}(s))} \right\} + \frac{\lambda_2(H_{20}(s))}{K(s)} - \frac{D_1(s)A(s)}{K(s)} \\
\quad - \int_s^\tau \frac{D_1(v)A(v)}{K(s)} \frac{\lambda_2(H_{20}(s))}{\gamma_2(H_{20}(v))} d\gamma_2(H_{20}(v)) \right\}, \\
C(t) = \int_0^t E \left[ Y_i(s) \frac{\partial X_i^0}{\partial \Theta} d\gamma_1(H_{10}(s), H_{20}(s)) \right] + E \left[ Y_i(s) \frac{\partial X_i^0}{\partial \Theta} d\gamma_2(H_{10}(s), H_{20}(s)) \right] \\
\quad - \frac{\partial \rho_0}{\partial \Theta} E \left[ Y_i(s) d\gamma_3(S(H_{10}(s), H_{20}(s))) \right], \\
\Delta = E \left\{ \frac{\partial^2 \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial \Theta'} \right\} + \int_0^\tau D_1(s) dC'(s) \\
\quad - \int_0^\tau D_1(s) A(s) \frac{d\mu'(s)}{\lambda_2(H_{20}(s))} \left[ \int_0^\tau D_1(v)A(v) d\gamma_2(H_{20}(v)) \right] \\
\quad + \int_0^\tau E \left\{ \frac{\partial^2 \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial H_2(U_{2i})} \frac{Y_{2i}(s)}{\lambda_2(H_{20}(s))} \right\} d\mu'(s) \right\}. \\
\]

**Theorem 1.** As \( n \to \infty \), we have
\[ |\hat{\Theta} - \Theta_0| \to 0, \quad \sup_{t \in [a, \tau]} |\hat{H}_1(t) - H_{10}(t)| \to 0, \quad \text{and} \quad \sup_{t \in [a, \tau]} |\hat{H}_2(t) - H_{20}(t)| \to 0 \]
in probability for any fixed \( a \in (0, \tau) \).

**Theorem 2.** As \( n \to \infty \), we have,
\[ \sqrt{n} \left( \hat{\Theta} - \Theta_0 \right) \to N(0, \Sigma^{-1} \Delta (\Sigma^{-1})'). \]

**Theorem 3.** As \( n \to \infty \), for any \( t \in (0, \tau) \), we have
\[ \sqrt{n} \left\{ \hat{H}_1(t) - H_{10}(t) \right\} \to N(0, \Sigma_1(t)), \]
where \( \Sigma_1(t) = E \left\{ -\zeta'(t) \Sigma^{-1} \frac{\partial \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta} \right\} + \int_0^\tau \omega_1(s, t) dM_1(s) - \int_0^\tau \omega_2(s, t) dM_2(s) \right\}^2, \]
\[ \omega_1(s, t) = I(s \leq t) \frac{\gamma_1(H_{10}(s))}{B(s) \gamma_1(H_{10}(t))} - \zeta'(t) \Sigma^{-1} D_1(s), \]
\[ \omega_2(s, t) = \zeta'(t) \Sigma^{-1} D_2(s) + I(s \leq t) \left\{ \frac{\lambda_2(H_{20}(s))}{B(s) K(s) \gamma_1(H_{10}(t))} \right\} + \frac{\lambda_2(H_{20}(s))}{K(s) \gamma_1(H_{10}(s))} \int_s^\tau A(v) \frac{\gamma_1(H_{10}(v))}{B(v) \gamma_2(H_{20}(v))} d\gamma_2(H_{20}(v)) \right\}, \]
and
\[
\zeta(t) = \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \frac{\gamma_1(H_{10}(s))dC(s)}{B(s)} - \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \left\{ \frac{\Lambda(s)\gamma_1(H_{10}(s))}{B(s)\lambda_2(H_{20}(s))} \right\} d\mu(s).
\]

**Theorem 4.** As \( n \to \infty \), for any \( t \in (0, \tau) \), we have
\[
\sqrt{n} \left( \tilde{H}_2(t) - H_{20}(t) \right) \to N(0, \Sigma_2(t)),
\]
where \( \Sigma_2(t) = \frac{1}{\lambda_2^2(H_{20}(t))} E \left\{ \int_0^t \frac{\lambda_2(H_{20}(s))}{\mathcal{K}(s)} dM_2(s) - \mu'(t) \Sigma^{-1} \partial \log \mathcal{L}_i(\Theta_0; H_{10}, H_{20}) \right\}
\]
\[
- \mu'(t) \Sigma^{-1} \int_0^\tau dM_i(s) - \mu'(t) \Sigma^{-1} \int_0^\tau dM_{2i}(s) \right\}^2.
\]

From Theorems 3 and 4, \( \tilde{H}_1(t) \) and \( \tilde{H}_2(t) \) converge to \( H_{10}(t) \) and \( H_{20}(t) \), respectively, at a rate of \( n^{-1/2} \). This result shows that we can estimate the nonparametric functions \( H_1(\cdot) \) and \( H_2(\cdot) \) with a parametric convergent rate. A similar conclusion, in which the transformation function can be estimated with \( n^{-1/2} \) rate of convergence, was also reached by Horowitz (1996), Chen (2002), Ye and Duan (1997) and Zhou, Lin and Johnson (2009).

As shown in Theorem 2, the asymptotic variance of \( \hat{\Theta} \) has the standard sandwich form \( \Sigma^{-1} \Delta (\Sigma^{-1})' \). However, the matrices \( \Sigma \) and \( \Delta \) are complicated analytic forms involving complicated computation. Therefore, a feasible computation approach is necessary to approximate the asymptotic variance of \( \hat{\Theta} \). In this article, a resampling scheme proposed by Jin, Ying, and Wei (2001) is used to approximate the asymptotic distribution of \( \hat{\Theta} \). The resampling algorithm proceeds as follows. First, we generate \( n \) exponential random variables \( \xi_i, i = 1, \cdots, n \) with mean 1 and variance 1. Fixing the data at their observed values, we solve the following \( \xi_i \)-weighted estimation equations and denote the solutions as \( \Theta^*, H^*_1(t) \) and \( H^*_2(t) \) for any \( t \in (0, \tau) \):

\[
\sum_{i=1}^n \xi_i \frac{\partial \mathcal{L}_i(\Theta; H_1, H_2)}{\partial \Theta} = 0,
\]
\[
\sum_{i=1}^n \xi_i \{ dN_2(t) + Y_2(t)d \log (1 - \Phi (H_2(t) - X_i'\alpha)) \} = 0,
\]
\[
\sum_{i=1}^n \xi_i \{ dN_1(t) + Y_1(t)d \log [S_p (H_1(t) - X_i'\beta, H_2(t) - X_i'\alpha)] \} = 0,
\]
where $H_1(0) = -\infty$ and $H_2(0) = -\infty$. The estimates $\Theta^*$, $H_1^*(t)$ and $H_2^*(t)$ can be obtained using the same iterative algorithm proposed in Section 2.3. Following the same lines as those in Jin et al. (2001), the validity of the proposed resampling method is established.

**Proposition.** The conditional distribution of $n^{1/2}(\Theta^* - \hat{\Theta})$ given the observed data converges almost surely to the asymptotic distribution of $n^{1/2}(\hat{\Theta} - \Theta_0)$.

Based on the Proposition, by repeatedly generating $\xi_1, \cdots, \xi_n$ many times, a large number of realizations of $\Theta^*$ can be obtained. The variance estimate of $\hat{\Theta}$ can then be approximated by the empirical variance of $\Theta^*$.

### 4 Simulation

In this section, simulations studies were conducted to assess the finite-sample performance of the proposed method by comparing it with the existing method. The existing approaches to analyze semicompeting risk data include (1) the bivariate location-shift regression model (BLSR) proposed by Lin et al. (1996); (2) the copula model; and (3) the combination of regression and copula model (CRC, Hsieh, Wang and Ding, 2008). The copula model is not yet ready for regression analysis, so we focus on the comparison of the proposed method with methods (1) and (3), termed by BLSR and CRC, respectively.

**Simulation 1.** The resulting estimates are expected to be reliable because our method does not require specification of a parametric form for the transformation function. Whether the added robustness is gained at the expense of reduced efficiency is also a purpose of this study. To investigate these issues, we examine the performance of the proposed method in comparison with the correct BLSR method, in which the transformation function is correctly specified (termed CBLSR), and the uncorrect BLSR method, in which the transformation function is misspecified (termed MBLSR). To compare with the BLSR estimator, we generate data with the sole binary covariate $X$ that took the value 1 for one half of the subjects and 0 for the other half, mimicking a binary treatment indicator. The simulation data are generated by the following model:

$$H_1(S) = \beta X + \epsilon_1, \quad H_2(T) = \alpha X + \epsilon_2,$$
where $H_1(t) = t$, $H_2(t) = \log t$, $\alpha = \beta = 1$, and $(\varepsilon_1, \varepsilon_2)'$ was a Gaussian vector with mean 0 and covariance matrix $\Sigma_\rho$, $\rho = 0.5$. $S$ is assumed to be the time to the non-terminal event and $T$ is the time to the terminal event. The censoring random variable $C$ is distributed uniformly on $(0, 20)$, so that about 15% of $T$ is censored by $C$ and about 15% of $S$ is censored by $C \land T$. A total of 200 simulations with a sample size of 200 are conducted for the simulation setting.

In the MBLSR, the transformations were misspecified as $H_1(t) = H_2(t) = t$. Table 1 presents the bias, the empirical standard deviation, the standardized bias (bias as a percent of the SD, termed as Sbias), and the root of mean square errors (RMSE) of the coefficient parameter estimators for $\beta$ and $\alpha$ based on the 200 simulations using the proposed method, the CBLSR and MBLSR methods. From Table 1, the following conclusions are drawn.

1. A useful rule of thumb to evaluate the biasedness is that biases do not have a substantial negative effect on inferences (e.g., by impairing the coverage of confidence intervals) unless the standardized bias exceeds 40% (Olsen and Schafer, 2001). By this rule, the proposed estimator and the CBLSR method are unbiased. In contrast, the MBLSR estimator is seriously biased and inefficient, especially for $\alpha$, which is the regression coefficient in the model where the transformation function is misspecified. The comparison of the CBLSR estimator with the MBLSR estimator shows that a correctly specified transformation function plays an important role in the performance of the BLSR methods. The misspecification of the transformation function can lead to the large biases and variances of the coefficient estimators.

2. By comparing the proposed estimates with the CBLSR estimates, we see that although the estimates from the proposed method have a larger bias than the CBLSR estimators, the proposed estimators are more efficient than those of the CBLSR method. As a result, the performances of the two methods are comparable in terms of mean square errors. The CBLSR estimator is a method that leaves the error distribution unspecified, whereas our estimator leaves the transformation function unspecified. Hence, correctly putting assumptions on the transformation functions or on the error distribution may not matter to the inference about the effect of covariates. However, it does matter to the association parameter because the CBLSR cannot directly provide the estimator of the association parameter while our method does.
Table 1 The bias, empirical standard deviation and root of mean square error (RMSE) of estimators based on 200 simulations.

<table>
<thead>
<tr>
<th>method</th>
<th>$\hat{\alpha}$ bias</th>
<th>se</th>
<th>Sbias</th>
<th>RMSE</th>
<th>$\hat{\beta}$ bias</th>
<th>se</th>
<th>Sbias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>0.0330</td>
<td>0.1089</td>
<td>0.3030</td>
<td>0.1138</td>
<td>0.0349</td>
<td>0.1047</td>
<td>0.3333</td>
<td>0.1104</td>
</tr>
<tr>
<td>CBLSR</td>
<td>0.0086</td>
<td>0.1173</td>
<td>0.0733</td>
<td>0.1176</td>
<td>-0.0029</td>
<td>0.1094</td>
<td>-0.0265</td>
<td>0.1094</td>
</tr>
<tr>
<td>MBLSR</td>
<td>1.3480</td>
<td>0.3506</td>
<td>3.8448</td>
<td>1.3928</td>
<td>0.1669</td>
<td>0.1318</td>
<td>1.2663</td>
<td>0.2127</td>
</tr>
<tr>
<td>Proposed</td>
<td>$\hat{\rho}$</td>
<td>0.0144</td>
<td>0.0387</td>
<td>0.3721</td>
<td>0.0413</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 1: Results of Simulation 1. (a) The averaged estimates of $H_1(t)$; (b) The averaged estimates of $H_2(t)$ (Solid —estimated and 95% confidence limit; dashed—true functions).

Figures 1(a) and 1(b) display the average of the estimated transformation functions and their pointwise 95% confidential intervals. From these figures we can see that the proposed method produces reasonable estimates of the transformation functions.

**Simulation 2.** Simulation 1 shows the misspecification of the transformation function will lead to a seriously biased estimator for the BLSR method. Our method requires the specification of the error distribution. A natural question is if the proposed method is sensitive to the error distribution. To investigate the issue, data
similar to those in Simulation 1 are generated except that the errors \((\varepsilon_1, \varepsilon_2)\)' jointly follow a Clayton copular model as

\[
Pr(\varepsilon_1 \geq x, \varepsilon_2 \geq y) = \phi_\gamma^{-1} [\phi_\gamma \{Pr(\varepsilon_1 \geq x)\} + \phi_\gamma \{Pr(\varepsilon_2 \geq y)\}],
\]

with \(\phi_\gamma(v) = (v^{-\gamma} - 1)/\gamma\), \(\gamma = 0.5\), and both the marginal distributions of \(\varepsilon_1 + 1\) and \(\varepsilon_2 + 1\) are chi-square distribution with one degree of freedom. Therefore, the assumption on the error distribution required by our method is not satisfied, but it follows the requirement of Hsieh, Wang and Ding (2008). A total of 200 simulations with a sample size of 200 are conducted for the simulation setting.

We are also interested in the comparison of the proposed method and the CRC method (Hsieh, Wang and Ding, 2008). Hence, for each simulated data, \(\beta\), \(\alpha\) and the association parameter are estimated using the proposed method, the CBLSR method, the MBLSR method, the CRC1 method and the CRC2 method. The CRC1 method is the CRC method with the transformation function correctly specified but the error distribution unspecified, and the CRC2 method is the CRC method with the error distribution correctly specified but the transformation function unspecified. The transformation functions are misspecified as \(H_1(t) = H_2(t) = t\) in the MBLSR method. Table 2 presents the bias, the empirical standard deviation, the standardized bias and the RMSE of \(\beta\), \(\alpha\) and the association parameter based on the 200 simulations. From Table 2, our estimator is slightly biased due to the misspecification of the error distribution, while the MBLSR estimator is seriously biased and inefficient. This result implies the estimation of the effect of covariates is driven more by the assumptions about the form of the transformation function than those about the error distribution. The conclusion is consistent with that founded by Lin and Zhou (2009).

From Table 2, the proposed method also shows a slightly larger bias and variance than the CRC method. This is not surprising because the error distribution does not follow our requirement, but follows the requirement of Hsieh, Wang and Ding (2008). In practice, the true error distribution can never be known, hence, the little loss of bias and efficiency seems acceptable.

| Table 2 | The bias, empirical standard deviation, the standardized bias and root of mean square error (RMSE) of estimators based on 200 simulations for Simulation 2. | 16 |
Simulation 3. In simulation 3, we consider the data with two-dimensional covariate vector, which is a combination of continuous and discrete covariates. The setting in Simulation 3 is similar to that in Simulation 1, except that the covariate \( Z = (Z_1, Z_2) \), \( H_1(t) = t^3 \) and \( H_2(t) = \Phi^{-1}(t/5) \), where \( Z_1 \) is generated uniformly over \([-2, 2]\), \( Z_2 \) is the treatment indicator in which \( n/2 \) subjects receive each of the two groups. The censoring random variable \( C \) is distributed uniformly on \((0, 20)\), so that about 15% of \( T \) is censored by \( C \) and about 5% of \( S \) is censored by \( C \land T \).

Lin et al. (1996) and Hsieh et al.’s (2008) methods are developed for low-dimension discrete covariate. To analyze the simulated data using Lin et al. and Hsieh et al. methods, the continuous covariates need to be grouped. It is well known that discretization may lead to information loss. To investigate the issue, we analyze the simulated data using the proposed method with the original covariates and the proposed method with the grouped covariates \( (Z_1^*, Z_2) \), where \( Z_1^* = I(Z_1 \geq 0) - I(Z_1 < 0) \). Table 3 presents the estimating results. Based on Table 3, even though the discretization can give the estimator less variance, the bias becomes much larger. As a result, the discretization leads to a much larger mean squares error, especially for the associated parameter and the coefficients of the discredited covariates.

Figure 2 displays the average of the estimated transformation functions and their empirical pointwise 95% confidential intervals. From these figures, the discretization also leads to biased estimators of the transformation functions, especially for the boundary. The proposed method with the original covariates produces reasonable estimates of the transformation functions.

Table 3 The bias, empirical standard deviation, the standardized bias and root of mean square error (RMSE) of estimators based on 200 simulations for Simulation 3.
<table>
<thead>
<tr>
<th>method</th>
<th>$\hat{\alpha}_1$ bias</th>
<th>$\hat{\alpha}_1$ se</th>
<th>Sbias</th>
<th>RMSE</th>
<th>$\hat{\alpha}_2$ bias</th>
<th>$\hat{\alpha}_2$ se</th>
<th>Sbias</th>
<th>RMSE</th>
</tr>
</thead>
<tbody>
<tr>
<td>Proposed</td>
<td>-0.0059</td>
<td>0.0653</td>
<td>-0.0904</td>
<td>0.0656</td>
<td>0.0531</td>
<td>0.1220</td>
<td>0.4352</td>
<td>0.1331</td>
</tr>
<tr>
<td>Discretized</td>
<td>-0.1417</td>
<td>0.0646</td>
<td>-2.1935</td>
<td>0.1557</td>
<td>-0.0867</td>
<td>0.1220</td>
<td>-0.7107</td>
<td>0.1497</td>
</tr>
<tr>
<td>Proposed</td>
<td>-0.0074</td>
<td>0.0608</td>
<td>-0.1217</td>
<td>0.0612</td>
<td>0.0390</td>
<td>0.1204</td>
<td>0.3239</td>
<td>0.1266</td>
</tr>
<tr>
<td>Discretized</td>
<td>-0.1378</td>
<td>0.0608</td>
<td>-2.2664</td>
<td>0.1506</td>
<td>-0.0946</td>
<td>0.1137</td>
<td>-0.8320</td>
<td>0.1479</td>
</tr>
<tr>
<td>Proposed</td>
<td>0.0187</td>
<td>0.0407</td>
<td>0.4595</td>
<td>0.0448</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Discretized</td>
<td>0.1413</td>
<td>0.0307</td>
<td>4.6026</td>
<td>0.1446</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

Figure 2: Results of Simulation 1. Top: the averaged estimates of $H_1(t)$ and $H_2(t)$ with the grouped covariates; Bottom: the averaged estimates of $H_1(t)$ and $H_2(t)$ with the original covariates (Solid —estimated and 95% confidence limit; dashed— true functions).
5 Example

Multiple myeloma is an incurable malignancy that originates in the antibody-secreting bone marrow plasma cells. Median survival is approximately 3-4 years, but the clinical course is highly variable and difficult to predict. Therefore, there is a need to better define patient-specific treatment strategies for the use of both standard and novel therapies. This dataset is about myeloma patients treated with a new agent (proteasome inhibitor bortezomib) or an active control drug (high-dose dexamethasone; Dex). A number of clinical and laboratory features provide prognostic information, including age, gender, proliferative index, as well as albumin and Myeloma score (expression of myeloma markers). Some of these factors relate to the patient’s status, whereas others reflect aspects of the tumor.

The data consists of 264 myeloma patient samples with 235 patients having complete information. Clinical responses were PGx_Days_To_Progression (OS) and time to death. OS was assessed from the date patients received their first dose of study drug, without regard to other subsequent therapies. Obviously, the overall survival and PGx_Days_To_Progression are the semicompeting outcomes.

The data is fitted by the following models:

\[ H_1(S) = X\beta + \varepsilon_1, \]
\[ H_2(T) = X\alpha + \varepsilon_2, \]

where \( S \) is the time to progression and \( T \) is survival time. The covariates \( X \) included treatment agent (Trt, 1=PS341, 0=Dex), gender (1=female), meanMyelScore (Score), Mayo_Clinic_ProliferativeIndex (Index), age and albumin. Here, we did not consider variable C_Reactive_Protein because nearly 30% of C_Reactive_Protein are missing.

The resulting estimates of the regression coefficients and associate parameter and their standard errors are listed in Table 4. The calculation of the standard errors was done via the resampling method described in Section 3 with 400 bootstrap samples. The choice of 400 sample size was determined by monitoring the stability of the standard errors.

Table 4 The resulting estimates for the myeloma data
<table>
<thead>
<tr>
<th></th>
<th>Estimator</th>
<th>SD</th>
<th>P-value</th>
<th></th>
<th>Estimator</th>
<th>SD</th>
<th>P-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Trt</td>
<td>0.185</td>
<td>0.203</td>
<td>0.363</td>
<td></td>
<td>0.511</td>
<td>0.223</td>
<td>0.022</td>
</tr>
<tr>
<td>Gender</td>
<td>0.064</td>
<td>0.160</td>
<td>0.689</td>
<td></td>
<td>-0.012</td>
<td>0.194</td>
<td>0.949</td>
</tr>
<tr>
<td>Score</td>
<td>-0.018</td>
<td>0.006</td>
<td>0.007</td>
<td></td>
<td>-0.004</td>
<td>0.008</td>
<td>0.603</td>
</tr>
<tr>
<td>Index</td>
<td>-0.439</td>
<td>0.190</td>
<td>0.021</td>
<td></td>
<td>-0.333</td>
<td>0.242</td>
<td>0.169</td>
</tr>
<tr>
<td>Age</td>
<td>0.018</td>
<td>0.014</td>
<td>0.189</td>
<td></td>
<td>0.033</td>
<td>0.014</td>
<td>0.015</td>
</tr>
<tr>
<td>Albumin</td>
<td>0.083</td>
<td>0.017</td>
<td>0.000</td>
<td></td>
<td>0.077</td>
<td>0.019</td>
<td>0.000</td>
</tr>
</tbody>
</table>

\[ \hat{\rho} = 0.386 \quad 0.077 \quad 0.000 \]

Figure 3: Estimate of H1 and H2. Solid — estimated; Points — 95% confidence limit.

Finally, we proposed a procedure to check validity of the assumed semiparametric transformation models (1.3) and (1.4). First, we randomly divided the data into five subsets with equal sizes. Four of the subsets are used as the training set and the remaining are used as validation set. Then, we fit a model using the training set. For each subject in the validation set, we predicted the subject’s event number up to time \( t \) by

\[
\hat{E}N_i(t) = -\int_0^t Y_i(t) d \log \left\{ S_{\hat{\rho}} \left( \hat{H}_1(t) - X_i'\hat{\beta}, \hat{H}_2(t) - X_i'\hat{\alpha} \right) \right\},
\]
and
\[ \hat{E}N_{2i}(t) = -\int_{0}^{t} Y_{2i}(t) d\log \left(1 - \Phi \left( \hat{H}_2(t) - X_i'\hat{\alpha} \right) \right). \]

We investigated the performance of the model by examining the prediction error
\[ PE_{1i} = \int_{0}^{\tau} \left( N_{i}(t) - \hat{E}N_{i}(t) \right) d \left\{ \sum_{k=1}^{n} N_{ki}(t) \right\}, \]
and
\[ PE_{2i} = \int_{0}^{\tau} \left( N_{2i}(t) - \hat{E}N_{2i}(t) \right) d \left\{ \sum_{k=1}^{n} N_{2ki}(t) \right\}. \]

Figures 4 and 5 plot the prediction error.

Figure 4: The prediction error for H1 versus the covariates.

6 Discussion

In the current paper, we propose semiparametric transformation models for semicompeting risk data. Our models allow the transformation function to be unknown but
the error distribution is specified to be normal. In this way, our model can provide direct estimators of the regression analysis and the association parameter. A simple algorithm is provided to estimate the transformation functions, and the proposed estimators are shown to be consistent and asymptotically normal. The simulation studies reveal that our method works very well compared to the existing method.

An important application of semicompeting risks approaches is assessing the surrogate endpoints. Surrogate end points are referred to as end points that can be used in lieu of other end points in the evaluation of treatments or other interventions. They are useful because they can be measured earlier, more conveniently, or more frequently than the end point of interest, which are referred to as the “true” or “final” end points (Ellenberg and Hamilton, 1989). In the surrogacy literature, $S$ is the surrogate endpoints and $T$ is the true endpoint. Before a surrogate end point can replace a final end point in the evaluation of an experimental treatment, it must be formally “validated”. Prentice (1989) proposed a formal definition of surrogate end points and outlined how potential surrogate end points could be validated. Prentice’s criteria are too stringent and are not straightforward to verify (Fleming et al., 1994).

Figure 5: The prediction error for H2 versus the covariates.
Freedmen et al. (1992) introduced the proportion explained, which is the proportion of the treatment effect that is mediated by the surrogate. Using the multivariate normal theory, $\varepsilon_2 = \rho \varepsilon_1 + \varepsilon^*$, where $\varepsilon^* \sim N(0, 1 - \rho^2)$ and is independent of $\varepsilon_1$. By coupling this with the models (1.3) and (1.4), we get

$$H_2(T) = \rho H_1(S) + X'(\alpha - \rho \beta) + \varepsilon^*. \tag{6.1}$$

Hence, by the definition given by Freedman, Graubard, and Schatzkin (1992), if $X$ is the indicator of treatment, the proportion of treatment effect (PTE) explained by the surrogate $S$ is $\rho \beta / \alpha$. This implies that the measure of surrogacy, or the association between $S$ and $T$ by the models (1.3) and (1.4), can be obtained. In contrast, this does not happen with the proportional hazards model (Lin, Fleming, and Degruttola, 1997) and the accelerated failure time model (Lin et al., 1996; Chang, 2000), which require an extra model to estimate PTE. Buyse and Molenberghs (1998) proposed replacing the proportion explained by two new measures. The first, termed the relative effect, is the ratio of the overall treatment effect on the true end point over that on the surrogate end point. The second is the individual level association between both end points, after accounting for the effect of treatment, referred to as adjusted association. Our model also provides the relative effect $RE = \beta / \alpha$. An RE value is useful only if the variance of $H_1(T)$ and $H_2(S)$ are the same (Ghosh, 2009). Obviously, in our model setting, the variance of $H_1(T)$ and $H_2(S)$ are the same, hence, $\beta / \alpha$ in our models provide a useful measure of surrogacy. In contrast, this does not happen with the bivariate location-shift model (Lin et al., 1996; Chang, 2000; Ghosh, 2009) because $S \leq T$, so the variance of the two random variables will generally not be the same.

References


**Acknowledgements**

Lin’s research is supported by the Chinese Natural Science Foundation (11071197), the National Natural Science Funds for Distinguished Young Scholar of China (No. 11125104) and Program for New Century Excellent Talents in University of China.

**7 Appendix: The proof of Theorems**

The consistency and asymptotic normality stated in theorems 1 and 2 are proved using similar arguments to those of Chen *et al.* (2002), so we only highlight the steps that are different.

**The proof of Theorems 1 and 2.**

*Step 1.* Using similar arguments to Step A1 of Chen *et al.* (2002), it can be shown that $d\{\hat{H}_2(\cdot, \Theta_0), H_{20}(\cdot)\} \to 0$ almost surely, where $\hat{H}_2(\cdot, \Theta)$ is the function implicitly
defined as the unique solution of (2.2) for fixed $\Theta$ and

\[ d(G_1, G_2) = \sup_{t \in [0, \tau]} \left| \exp(G_1(t)) - \exp(G_2(t)) \right|. \]

Now we show that $D\{ \hat{H}_1(\cdot, \Theta_0), H_{20}(\cdot) \} \to 0$ almost surely, where $\hat{H}_1(\cdot, \Theta_0) \in \mathcal{H}_1$ is the function implicitly defined as the unique solution of (2.3) with $\Theta = \Theta_0$ and $H_2(\cdot) = \hat{H}_2(\cdot, \Theta_0)$, and

\[ D(G_1, G_2) = \sup_{t \in [0, \tau]} \left| E \left[ \log \left\{ S_{\rho_0} \left( G_1(t \wedge U_{1i}) - X_i' \beta_0, H_{20}(t \wedge U_{1i}) - X_i' \alpha_0 \right) \right\} \right. \right. \]
\[ - \log \left\{ S_{\rho_0} \left( G_2(t \wedge U_{1i}) - X_i' \beta_0, H_{20}(t \wedge U_{1i}) - X_i' \alpha_0 \right) \right\} \right] \right|, \]

for any two nondecreasing functions $G_1$ and $G_2$ on $[0, \tau]$ such that $G_1(0) = G_2(0) = -\infty$. Denote

$\mathcal{H}_1 = \{ H_1 : H_1$ is nondecreasing step functions on $[0, \tau]$ with $H_1(0) = -\infty \}

and with jumps only at the observed failure times $t_1, \cdots, t_M \}$, and $A$ a mapping defined by

\[ A(H_1)(t) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \left[ dN_i(t) + Y_i(t) d \log \left\{ S_{\rho_0} \left( H_1(t) - X_i' \beta_0, \hat{H}_2(t, \Theta_0) - X_i' \alpha_0 \right) \right\} \right], \]

where $\rho_0$, $\beta_0$ and $\alpha_0$ are the true values of $\rho$, $\beta$ and $\alpha$. For an arbitrary but fixed $\epsilon > 0$, consider $G_1$ and $G_2$ such that $D(G_1, G_2) > \epsilon$, then there exists a $t^* \in [0, \tau]$ such that

\[ \left| E \left[ \log \left\{ S_{\rho_0} \left( G_1(t^* \wedge U_{1i}) - X_i' \beta_0, H_{20}(t^* \wedge U_{1i}) - X_i' \alpha_0 \right) \right\} \right. \right. \]
\[ - \log \left\{ S_{\rho_0} \left( G_2(t^* \wedge U_{1i}) - X_i' \beta_0, H_{20}(t^* \wedge U_{1i}) - X_i' \alpha_0 \right) \right\} \right] \geq \epsilon/2. \]
Hence, coupling with \(d\{\hat{H}_2(\cdot, \Theta_0), H_{20}(\cdot)\} \to 0\) almost surely, we have

\[
\sup_{t \in [0, \tau]} \left| A(G_1)(t) - A(G_2)(t) \right| = \frac{1}{n} \sup_{t \in [0, \tau]} \left| \sum_{i=1}^{n} \left[ \log \left\{ S_{\rho_0} \left( G_1(t \land U_{1i}) - X'_i \beta_0, \hat{H}_2(t \land U_{1i}, \Theta_0) - X'_i \alpha_0 \right) \right\} 
- \log \left\{ S_{\rho_0} \left( G_2(t \land U_{1i}) - X'_i \beta_0, \hat{H}_2(t \land U_{1i}, \Theta_0) - X'_i \alpha_0 \right) \right\} \right| \right|
\ge \frac{1}{n} \left| \sum_{i=1}^{n} \left[ \log \left\{ S_{\rho_0} \left( G_1(t^* \land U_{1i}) - X'_i \beta_0, \hat{H}_2(t^* \land U_{1i}, \Theta_0) - X'_i \alpha_0 \right) \right\} 
- \log \left\{ S_{\rho_0} \left( G_2(t^* \land U_{1i}) - X'_i \beta_0, \hat{H}_2(t^* \land U_{1i}, \Theta_0) - X'_i \alpha_0 \right) \right\} \right| \right|
\ge \epsilon/2
\]

when \(n\) is large enough. Choose \(H_{10}^* \in \mathcal{H}_1\) such that \(H_{10}^*(t_i) = H_{10}(t_i)\) for \(i = 1, \cdots, M\). The law of large numbers, the continuity of \(H_{10}\) and \(d\{\hat{H}_2(\cdot, \Theta_0), H_{20}(\cdot)\} \to 0\) imply that \(\sup \{A(H_{10}^*)(t) : t \in [0, \tau]\} \to 0\) almost surely. It follows that

\[
\sup_{t \in [0, \tau]} \left| A(H_{10}^*)(t) - A(\hat{H}_1(\cdot, \Theta_0))(t) \right| \to 0
\]

almost surely because, by definition of \(\hat{H}_1(\cdot, \Theta_0), A(\hat{H}_1(\cdot, \Theta_0))(t) = 0\) for all \(t \in [0, \tau]\). Then, with probability 1, \(\hat{H}_1(\cdot, \Theta_0)\) is in the neighborhood of \(H_{10}^*\) of radius \(\epsilon\) under the metric \(D(\cdot, \cdot)\). Therefore, \(D(\hat{H}_1(\cdot, \Theta_0), H_{10}) \to 0\) almost surely, since \(\epsilon > 0\) can be arbitrarily small and \(\hat{H}_1(\cdot, \Theta_0)\) and \(H_{10}\) are monotone.

**Step 2.** Constructing the expressions of \(\hat{H}_2(t; \Theta_0)\) and \(\hat{H}_1(t; \Theta_0)\). Let \(a > 0\) and \(b\) be fixed finite numbers and define

\[
\mathcal{K}_1(t) = \int_a^t E[Y_2(s)\lambda_{2i}^{(1)}(H_{20}(s))]dH_{20}(s), \quad \Gamma_1(x) = \int_b^x \gamma_1(s)ds, \quad \Lambda_2(x) = \int_b^x \lambda_2(s)ds,
\]

for \(t > 0\) and \(x \in (-\infty, \infty)\). We choose finite \(a > 0\) and \(b\) as the lower limits of the integration to ensure that the integrals are finite. Similarly to Step A2 in Chen et al. (2002), we have, uniformly for \(t \in [0, \tau]\),

\[
\Lambda_2(\hat{H}_2(t; \Theta_0)) - \Lambda_2(H_{20}(t)) = \frac{1}{n} \sum_{i=1}^{n} \int_0^t \lambda_2(H_{20}(s)) \frac{dM_{2i}(s)}{\mathcal{K}(s)} + o_p(n^{-1/2}). \quad (7.1)
\]

Now we consider the representative of \(\hat{H}_1(t; \Theta_0)\). Denote

\[
B_1(t) = \int_a^t E[Y_1(s)d\gamma_{1i}(H_{10}(s), H_{20}(s))], \quad \Gamma_2(x) = \int_b^x \gamma_2(s)ds,
\]

27
\[ A_1(t) = \int_0^t E[Y_i(s)d\gamma_1(H_{10}(s), H_{20}(s))], \quad A(t) = E[Y_i(t)\gamma_2(H_{10}(t), H_{20}(t))], \]

It is easy to see that \(d\gamma_1\{H_{10}(t)\} = \frac{\gamma_1\{H_{10}(t)\}}{B(t)}dB_1(t)\) and \(d\gamma_2\{H_{20}(t)\} = \frac{\gamma_2\{H_{20}(t)\}}{A(t)}dA_1(t)\) and write

\[
\frac{1}{n} \sum_{i=1}^n M_i(t) = -\frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) \left\{ \log \left[ S_{\rho_o} \left( \hat{H}_1(s, \Theta_0) - X_i'\beta_0, \hat{H}_2(s, \Theta_0) - X_i'\alpha_0 \right) \right] - \log \left[ S_{\rho_o} \left( H_{10}(s) - X_i'\beta_0, H_{20}(s) - X_i'\alpha_0 \right) \right] \right\} ds
\]

\[
= \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) \left\{ \frac{\gamma_1(H_{10}(s), H_{20}(s))}{\gamma_1(H_{10}(s))} \left( \Gamma_1(\hat{H}_1(s; \Theta_0)) - \Gamma_1(H_{10}(s)) \right) \right\} ds
\]

\[
+ \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) \left\{ \frac{\gamma_2(H_{10}(s), H_{20}(s))}{\gamma_2(H_{20}(s))} \left( \Gamma_2(\hat{H}_2(s; \Theta_0)) - \Gamma_2(H_{20}(s)) \right) \right\} ds + o_p(n^{-1/2})
\]

\[
= \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) \left\{ \frac{\gamma_1(H_{10}(s), H_{20}(s))}{\gamma_1(H_{10}(s))} \left( \Gamma_1(\hat{H}_1(s; \Theta_0)) - \Gamma_1(H_{10}(s)) \right) \right\} ds
\]

\[
+ \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) \left\{ \frac{\gamma_2(H_{10}(s), H_{20}(s))}{\gamma_2(H_{20}(s))} \left( \Gamma_2(\hat{H}_2(s; \Theta_0)) - \Gamma_2(H_{20}(s)) \right) \right\} ds + o_p(n^{-1/2})
\]

\[
= \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) \frac{\gamma_1(H_{10}(s), H_{20}(s))}{\gamma_1(H_{10}(s))} d \left( \Gamma_1(\hat{H}_1(s; \Theta_0)) - \Gamma_1(H_{10}(s)) \right)
\]

\[
+ \frac{1}{n} \sum_{i=1}^n \int_0^t Y_i(s) \frac{\gamma_2(H_{10}(s), H_{20}(s))}{\gamma_2(H_{20}(s))} d \left( \Gamma_2(\hat{H}_2(s; \Theta_0)) - \Gamma_2(H_{20}(s)) \right) + o_p(n^{-1/2})
\]

\[
= \int_0^t \frac{B(s)}{\gamma_1(H_{10}(s))} d \left( \Gamma_1(\hat{H}_1(s; \Theta_0)) - \Gamma_1(H_{10}(s)) \right)
\]

\[
+ \int_0^t \frac{A(s)}{\gamma_2(H_{20}(s))} d \left( \Gamma_2(\hat{H}_2(s; \Theta_0)) - \Gamma_2(H_{20}(s)) \right) + o_p(n^{-1/2}). \quad (7.2)
\]
Denote \( \Upsilon(t) = \Gamma_2(\Lambda_2^{-1}(t)) \). By (7.1), uniformly for \( t \in [0, \tau] \), we have

\[
\Gamma_2(\hat{H}_2(t; \Theta_0)) - \Gamma_2(H_2(t)) = \Upsilon(\Lambda_2(\hat{H}_2(t; \Theta_0))) - \Upsilon(\Lambda_2(H_2(t))) \\
= \frac{\gamma_2(H_2(t))}{n \lambda_2(H_2(t))} \sum_{i=1}^n \int_0^t \frac{\lambda_2(H_2(s))}{K(s)} dM_{2i}(s) + o_p(n^{-1/2}).
\]

(7.3)

Substituting it into (7.2), we get

\[
\Gamma_1(\hat{H}_1(t; \Theta_0)) - \Gamma_1(H_0(t)) = \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\gamma_1(H_{10}(s))}{B(s)} dM_i(s) \\
- \int_0^t \frac{A(s) \gamma_1(H_{10}(s))}{B(s)} \left( \gamma_2(H_{20}(s)) \sum_{i=1}^n \int_0^s \frac{\lambda_2(H_{20}(u))}{K(u)} dM_{2i}(u) \right) + o_p(n^{-1/2}) \\
= \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{\gamma_1(H_{10}(s))}{B(s)} dM_i(s) - \frac{1}{n} \sum_{i=1}^n \int_0^t \frac{A(s) \gamma_1(H_{10}(s))}{B(s)} \frac{\lambda_2(H_{20}(s))}{K(s)} dM_{2i}(s) \\
- \int_0^t \frac{A(s) \gamma_1(H_{10}(s))}{B(s)} \left( \frac{1}{n} \sum_{i=1}^n \int_0^s \frac{\lambda_2(H_{20}(u))}{K(u)} dM_{2i}(u) \right) + o_p(n^{-1/2}) \\
= \frac{1}{n} \sum_{i=1}^t \int_0^t \frac{\gamma_1(H_{10}(s))}{B(s)} dM_i(s) - \frac{1}{n} \sum_{i=1}^t \int_0^t \frac{A(s) \gamma_1(H_{10}(s))}{B(s)} \frac{\lambda_2(H_{20}(s))}{K(s)} dM_{2i}(s) \\
+ \frac{\lambda_2(H_{20}(s))}{K(s)} \int_0^t \frac{A(s) \gamma_1(H_{10}(s))}{B(s)} \left( \frac{\lambda_2(H_{20}(s))}{K(s)} \right) dM_{2i}(s) + o_p(n^{-1/2}) \\
\approx \frac{1}{n} \sum_{i=1}^t \int_0^t \Psi_1(s) dM_i(s) - \frac{1}{n} \sum_{i=1}^t \int_0^t \Psi_2(s, t) dM_{2i}(s) + o_p(n^{-1/2}),
\]

(7.4)

uniformly over \( t \in [0, \tau] \).

**Step 3.** Denote \( \mathcal{L}(\Theta; H_1, H_2) = \prod_{i=1}^{n} \mathcal{L}_i(\Theta; H_1, H_2) \) and \( U(\Theta; H_1, H_2) = \frac{\partial \log \mathcal{L}(\Theta; H_1, H_2)}{\partial \Theta} \) at \( \Theta = \Theta_0 \). By differentiating both side of (2.2) with \( H_2(t) \) replaced by \( \hat{H}_2(t; \Theta) \), respect to \( \Theta \), we obtain the identity

\[
\sum_{i=1}^n \int_0^t Y_{2i}(s) d \left\{ \lambda_2(\hat{H}_2(s; \Theta)) \left( \frac{\partial \hat{H}_2(s; \Theta)}{\partial \Theta} - \frac{\partial X'\alpha}{\partial \Theta} \right) \right\} = 0.
\]

(7.5)

Similarly to Step 2, we have that

\[
\frac{\partial \hat{H}_2(t; \Theta_0)}{\partial \Theta} = \frac{\mu(t)}{\lambda_2(H_{20}(t))} + o_p(1),
\]

(7.6)

where \( \mu(t) \) is defined in Section 3.
On the other hand, by differentiating both side of (2.3) with \( \hat{H}_1(t; \Theta) \) and \( \hat{H}_2(t; \Theta) \), respect to \( \Theta \), we obtain the identity

\[
\sum_{i=1}^{n} Y_i(t) \left\{ \gamma_{11}(\hat{H}_1(t; \Theta), \hat{H}_2(t; \Theta)) \left[ \frac{\partial \hat{H}_1(t; \Theta)}{\partial \Theta} - \frac{\partial X_i'\beta}{\partial \Theta} \right] + \gamma_{21}(\hat{H}_1(t; \Theta), \hat{H}_2(t; \Theta)) \left[ \frac{\partial \hat{H}_2(t; \Theta)}{\partial \Theta} - \frac{\partial X_i'\alpha}{\partial \Theta} \right] \right\} = 0,
\]

where \( \gamma_{3i}(x, y) = \frac{S_x(x - X_i'\beta)}{S_x(x - X_i'\beta - X_i'\alpha)} \) and \( \hat{S}_\rho(x, y) = \frac{\partial S_x(x, y)}{\partial \rho} \). Similarly to Step 2, we have that

\[
\frac{\partial \hat{H}_1(t, \Theta)}{\partial \Theta} \gamma_1(H_{10}(t)) + \int_0^t \frac{A(s)\gamma_1(H_{10}(s))}{B(s)\gamma_2(H_{20}(s))} d\left( \frac{\partial \hat{H}_3(s, \Theta)}{\partial \Theta} \right) = \int_0^t \gamma_1(H_{10}(s)) \frac{dC(s)}{B(s)},
\]

for \( \Theta = \Theta_0 \), where \( C(s) \) is defined in Section 3. Then, substituting (7.6) into it, we get

\[
\frac{\partial \hat{H}_1(t, \Theta)}{\partial \Theta} = \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \frac{\gamma_1(H_{10}(s))}{B(s)} dC(s) - \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \frac{A(s)\gamma_1(H_{10}(s))}{B(s)\lambda_2(H_{20}(s))} d\mu(s)
\]

\[
- \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \frac{\gamma_1(H_{10}(s))A(s)\mu(s)}{\lambda_2(H_{20}(s))B(s)} \left( \frac{d\gamma_2(H_{20}(s))}{\gamma_2(H_{20}(s))} - \frac{d\lambda_2(H_{20}(s))}{\lambda_2(H_{20}(s))} \right) d\mu(s)
\]

\[
= \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \frac{\gamma_1(H_{10}(s))}{B(s)} dC(s) - \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \frac{A(s)\gamma_1(H_{10}(s))}{B(s)\lambda_2(H_{20}(s))} d\mu(s)
\]

\[
- \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \int_s^t \frac{\gamma_1(H_{10}(v))A(v)}{\lambda_2(H_{20}(v))B(v)} \left( \frac{dA_1(v)}{A(v)} - \frac{d\lambda_2(H_{20}(v))}{\lambda_2(H_{20}(v))} \right) d\mu(s)
\]

\[
= \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \frac{\gamma_1(H_{10}(s))}{B(s)} dC(s) - \frac{1}{\gamma_1(H_{10}(t))} \int_0^t \left\{ \frac{A(s)\gamma_1(H_{10}(s))}{B(s)\lambda_2(H_{20}(s))} \right\} d\mu(s) \quad \text{for } \Theta = \Theta_0,
\]

It follows from the law of large numbers that

\[
V(\Theta_0) = \frac{1}{n} \frac{\partial^2 \log L(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial \Theta'} + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial \hat{H}_1(U_{1i}; \Theta_0)}
\]

\[
+ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial \hat{H}_2(U_{2i}; \Theta_0)}
\]

\[
= \Sigma + o_p(1),
\]
where \( L_i(\boldsymbol{\Theta}; H_1, H_2) \) is the contribution of subject \( i \) to the likelihood function \( L(\boldsymbol{\Theta}; H_1, H_2) \).

**Step 4.** In the step, we show the asymptotic normality of \( U(\Theta_0; \hat{H}_1(\cdot, \Theta_0), \hat{H}_2(\cdot, \Theta_0)) \). Using the results of Steps 1 and 2 and some empirical process approximation techniques, we can write

\[
\frac{1}{n} U(\Theta_0; \hat{H}_1(\cdot, \Theta_0), \hat{H}_2(\cdot, \Theta_0)) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta} + \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial H_1(U_{1i})} \left( \hat{H}_1(U_{1i}; \Theta_0) - H_{10}(U_{1i}) \right) \\
+ \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial H_2(U_{2i})} \left( \hat{H}_2(U_{2i}; \Theta_0) - H_{20}(U_{2i}) \right) + o_p(n^{-1/2})
\]

Substituting (7.4) and (7.3) into it, we get

\[
\frac{1}{n} U(\Theta_0; \hat{H}_1(\cdot, \Theta_0), \hat{H}_2(\cdot, \Theta_0)) = \frac{1}{n} \sum_{i=1}^{n} \frac{\partial \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta} + \frac{1}{n} \sum_{j=1}^{n} \frac{\partial^2 \log L_j(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial H_1(U_{1j})} \frac{1}{\gamma_1(H_{10}(U_{1j}))} \\
\times \left( \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{U_{1j}} \Psi_1(s) dM_i(s) - \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{U_{1j}} \Psi_2(s, U_{1j}) dM_2(s) \right) \\
+ \frac{1}{n} \sum_{j=1}^{n} \frac{\partial^2 \log L_j(\Theta_0; H_{10}, H_{20})}{\partial \Theta \partial H_2(U_{2j})} \frac{1}{\lambda_2(H_{20}(U_{2j}))} \left( \frac{1}{n} \sum_{i=1}^{n} \int_{0}^{U_{2j}} \lambda_2(H_{20}(s)) dM_2(s) \right) + o_p(n^{-1/2})
\]

\[
= \frac{1}{n} \sum_{i=1}^{n} \left\{ \frac{\partial \log L_i(\Theta_0; H_{10}, H_{20})}{\partial \Theta} + \int_{0}^{T} D_1(s) dM_i(s) + \int_{0}^{T} D_2(s) dM_2(s) \right\} + o_p(n^{-1/2})
\]

where \( D_1(s) \) and \( D_2(s) \) are defined in Section 3. It then follows that

\[
\frac{1}{\sqrt{n}} U(\Theta_0; \hat{H}_1(\cdot, \Theta_0), \hat{H}_2(\cdot, \Theta_0)) \rightarrow N(0, \Delta),
\]

where \( \Delta \) is defined in Section 3.
The rest of the proof essentially proceeds along the lines of Chen et al. (2002) and is omitted here.

The proof of Theorem 3.

By Taylor series expansions, (7.3), (7.6) and Theorem 2, we get,

\[ \Lambda_2(\hat{H}_2(t; \Theta)) - \Lambda_2(H_{20}(t)) = \Lambda_2(\tilde{H}_2(t; \Theta)) - \Lambda_2(\tilde{H}_2(t; \Theta_0)) + \Lambda_2(\hat{H}_2(t; \Theta_0)) - \Lambda_2(H_{20}(t)) \]

\[ = \lambda_2(H_{20}(t)) \frac{\partial \hat{H}_2(t; \Theta_0)}{\partial \Theta} (\Theta - \Theta_0) + \Lambda_2(\tilde{H}_2(t; \Theta_0)) - \Lambda_2(H_{20}(t)) + o_p(n^{-1/2}) \]

\[ = \frac{1}{n} \sum_{i=1}^{n} \left\{ \int_0^t \frac{\lambda_2(H_{20}(s))}{K_i(s)} dM_i(s) - \mu'(t) \Sigma^{-1} \frac{\partial}{\partial \Theta} \log L_i(\Theta_0; H_{10}, H_{20}) \right. \]

\[ \left. - \mu'(t) \Sigma^{-1} \int_0^t D_1(s) dM_i(s) - \mu'(t) \Sigma^{-1} \int_0^T D_2(s) dM_2(s) \right\} + o_p(n^{-1/2}). \]

Then Theorem 3 follows.

The proof of Theorem 4.

By Taylor series expansions, we get,

\[ \Gamma_1(\hat{H}_1(t)) - \Gamma_1(H_{10}(t)) \]

\[ = \Gamma_1(\tilde{H}_1(t; \Theta)) - \Gamma_1(\tilde{H}_1(t; \Theta_0)) + \Gamma_1(\hat{H}_1(t; \Theta_0)) - \Gamma_1(H_{10}(t)) \]

\[ = \gamma_1(H_{0}(t)) \frac{\partial \hat{H}_1(t; \Theta_0)}{\partial \Theta} (\Theta - \Theta_0) + \Gamma_1(\tilde{H}_1(t; \Theta_0)) - \Gamma_1(H_{0}(t)). \]

Then Theorem 4 follows from Theorem 2, (7.4) and (7.7).